# Introduction

This document describes the project plan for creating a set of HDF tools use-cases or examples. In the past, users have asked us many questions regarding using HDF5 tools and viewing file content in HDF5 files. Our current document does not handle those questions well. We have to rely on our help desk or the tools team to answer those questions repeatedly. Having a set of good examples or use-cases will be very helpful for our users and us. Simple use-cases will help our users, especially new users, to easily understand and adopt the HDF5 technologies. This set of use-cases will be strategically important for the company to expand the use of the HDF technologies.

We currently do not have a set of good use-cases for our HDF5 tools. As a result, users have to spend a lot of time trying to figure out solutions even though many of the solutions are already available. Others have to ask our help desk or the forum for assistance. A good set of use-cases will definitely help our users and our team to save a lot of unnecessary effort.

The purpose of this document is to identify the tasks for this work and layout the implementation details. The goal of this work is to make a set of simple use-cases for HDF5 tools that will help users easily understand and adopt HDF5.

# Stakeholders

We have three basic types of HDF5 data users: data producers, data consumers, and applications developers. Our stakeholders should include all the three types of users:

* Data consumers are the ones who use the data. The data consumers, especially the ones who are new to HDF5, are mostly interested in the data content. Simple use cases of showing data content will be useful for them.
* Application developers are the ones who build applications using HDF5. Those users understand not only HDF5 files but also HDF5 library API functions. They may want to know more details of file content, such as compression and storage layout. They will look for advanced use cases.
* Data producer are the ones who create HDF5 file in their production line. The files can be produced in their scripts or by their applications. These users may want to know the details of files for verification purpose. Use cases such as diff files or objects will be very useful to them.

# Scope

This project will produce a set of examples on how to work with HDF5 files with HDF5 tools. The set of examples includes:

* Creating a new file with a simple group structure
* Browsing file objects
* Reading data content
* Viewing metadata (attributes, datatypes, dimensions, storage layouts, etc.)
* Exporting data to text or binary files
* Importing text or binary data
* Repacking a file
* Using h5diff

The deliverables include the set of examples and related documents. The final product will be reviewed by the technical team and our major paying customers before it is published at the HDF group website.

This work only provides limited users cases or examples. The set of examples can help users to understand HDF5 files but it cannot serve as the HDF5 users’ guide nor can it be used to explore the full features of HDF5. No API functions will be used. The examples are all high level use of the HDF5 tools and will not help users how to understand the library functions.

# Product Description

## High-level product requirements

The basic requirements for the set of examples include:

* Easy-to-use: the examples should be simple and straightforward
* A fair coverage of main use-cases: viewing, editing, and creating file and data content
* Well documented:

## Break-down descriptions

In the following sections we describe the user’s cases in more details. Below is a list the definitions/items that are applied to all the user’s cases.

* Stakeholders and Interests: all users, especially those who are new to HDF5.
* Preconditions and guarantees: users must have some basic knowledge of HDF5, such as groups and files. Two things will be
* Main Success Scenario: users can easily follow the instructions and produce the same results as documented.

### Creating a new file with a simple group structure

* Goal in Context: a user wants to create a HDF5 file with some simple group structure. There are several options the user can do: writing a program by using the HDF library API functions; using command-line tools, and using the HDF-Java library or HDFView. Our goal is to show the simplest and easiest ways to create a file so that any users, even those who do not know how to program in HDF5 library API functions, can do it quickly.
* Scope: h5mkgrp and HDFView will be presented to create a HDF5 file with a group at the root, a subgroup under another group.
* Extensions: Following the similar steps, users should be able to create different group structures. However, creating datasets are not covered by these user’s cases.

### Browsing file objects

* Goal in Context: In most of the cases, the first thing users want to know is what things are in the file. When a file contains large datasets, simply dumping the file content is not a good way to figure out what objects are in the file. The goal is to show users how to browse HDF5 object in a file.
* Scope: three tools, h5ls, “h5dump –H”, and HDFView will be presented to show objects in a file. Showing content of datasets is not included in this user’s case

### Reading data content

* Goal in Context: viewing data content is what most users would like to do. The goal of this user’s case is to show different ways of looking data values.
* Scope: h5dump, h5ls, and HDFView will be used to show data content. We will also show data content in table or image in HDFView. Advanced data selection will not be included.

### Viewing metadata

* Goal in Context: The goal of this user’s case is to show metadata information that includes attributes, datatype, dataspace, and storage layout.
* Scope: h5dump and HDFView will be used to show metadata.

### Exporting data to text or binary file

* Goal in Context: export data values to a text file or a binary file.
* Scope: h5dump and HDFView will be used. Three types of datasets will be included: 2D floats, 2D image, and 1D compound.
* Extensions: Following the similar steps, users should be able to export data of other types.

### Importing text or binary data

* Goal in Context: importing to HDF5 is very important to users but can be complicated for complex datasets. These features are not well document. The goal of this user’s case is to demonstrate how the importing works in HDF5 tools.
* Scope: h5import and HDFView will be used. Two types of datasets will be included: 2D floats and 1D compound.
* Extensions: Following the similar steps, users should be able to import data of other types.

### Repacking a file

Most of our users do not know the value of h5repack tool. h5repack is very useful for many cases such as removing unreachable space in files and changing storage layout.

* Goal in Context: h5repack is a very useful tool for many things such as removing unreachable space in files and changing storage layout. The goal is to demonstrate the basic features of h5repack
* Scope: two examples will be included: remove unusable spaces and changing storage layout.
* Extensions: Following the similar steps, users should be able to explore the full features of h5repack.

### Using h5diff

* Goal in Context: h5diff is one of the most used tools and the most complex one. The goal is to show users how to use h5diff.
* Scope: two examples will be included: diff specific datasets and diff an entire file

# Risk management

Since the work is relative small and the work is straightforward, the risk is very low. The main risks would be that the staff resources are taken away from the project before the project is completed or the original estimation of the work for each task is too low. We will use the following approaches to reduce the risks.

* Agile approach: prioritize and deliver products (use examples) in an incremental order. The use-cases will be developed, documented, reviewed, and delivered piece by piece.
* Conservative estimation of work: try to be more convective on work estimation. We will try to include all the work such as discussion and review in the estimation.

# Planning

The tasks, responsibilities, estimated work, proposed schedule are summarized in the following table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task | Estimated work | Starting date | Ending date | Staff | Monitor |
| Project planning | 6 | 2/20/2013 | 3/1/2013 | Peter | Kent or Elena? |
| Review and revise the project plan | 2 | 3/13/2013 | 3/14/2013 | Mark, Peter | Kent or Elena? |
| Approve project plan | 2 | 3/14/2013 | 3/20/2013 | Elena | Kent or Elena? |
| Use-case #1: creating a new file | 2 | 3/20/2013 | 3/21/2013 | Peter | Kent or Elena? |
| Review and document use-case #1 | 2 | 3/21/2013 | 3/22/2013 | Mark | Kent or Elena? |
| Use-case #2: Browsing file objects | 3 | 3/22/2013 | 3/23/2013 | Peter | Kent or Elena? |
| Review and document use-case #2 | 2 | 3/23/2013 | 3/24/2013 | Mark | Kent or Elena? |
| Use-case #3: Reading data content | 4 | 3/24/2013 | 3/27/2013 | Peter | Kent or Elena? |
| Review and document use-case #3 | 2 | 3/27/2013 | 3/28/2013 | Mark | Kent or Elena? |
| Use-case #4: Viewing metadata | 4 | 3/28/2013 | 3/29/2013 | Peter | Kent or Elena? |
| Review and document use-case #4 | 2 | 3/29/2013 | 3/30/2013 | Mark | Kent or Elena? |
| Use-case #5: Exporting data | 2 | 3/30/2013 | 3/31/2013 | Peter | Kent or Elena? |
| Review and document use-case #5 | 2 | 3/31/2013 | 4/3/2013 | Mark | Kent or Elena? |
| Use-case #6: Importing data | 4 | 4/3/2013 | 4/4/2013 | Peter | Kent or Elena? |
| Review and document use-case #6 | 2 | 4/4/2013 | 4/5/2013 | Mark | Kent or Elena? |
| Use-case #7: Repacking a file | 2 | 4/5/2013 | 4/6/2013 | Peter | Kent or Elena? |
| Review and document use-case #7 | 2 | 4/6/2013 | 4/7/2013 | Mark | Kent or Elena? |
| Use-case #8: Using h5diff | 4 | 4/7/2013 | 4/10/2013 | Peter | Kent or Elena? |
| Review and document use-case #8 | 2 | 4/10/2013 | 4/11/2013 | Mark | Kent or Elena? |
| Total | 51 |  |  |  | Kent or Elena? |